Int J Life Cycle Assess (2014) 19:377–386
DOI 10.1007/s11367-013-0645-6

LIFE CYCLE SUSTAINABILITY ASSESSMENT

Impact of normalisation, elicitation technique and background information on panel weighting results in life cycle assessment

Tanja Myllyviita · Pekka Leskinen · Jyri Seppälä

Received: 16 October 2012 / Accepted: 13 August 2013 / Published online: 27 August 2013

Abstract
Purpose Weighting in Life Cycle Assessment (LCA) is a much-debated topic. Various tools have been used for weighting in LCA, Multi-Criteria Decision Analysis (MCDA) being one of the most common. However, it has not been thoroughly assessed how weight elicitation techniques of MCDA with different scales (interval and ratio) along with external and internal normalisation affect weighting and subsequent results. The aim of this survey is to compare different techniques in an illustrative example in the building sector.
Methods A panel of Nordic LCA experts accomplished six weighting exercises. The different weight elicitation techniques are SWING which is based on the interval scale; Simple Multi-Attribute Rating Technique (SMART) and Analytic Hierarchy Process (AHP) which is based on the ratio scale. Information on the case study was provided for the panellists, along with characterised or normalised impact assessment scores. However, in the first weighting exercise, the panellists were not provided with any scores or background information, but they had to complete the weighting at a more general level. With the weights provided by the panel, the environmental impacts of three alternative house types were aggregated. The calculations were based on three well-grounded aggregation rules, which are commonly used in the field of LCA or decision analysis.
Results and discussion In the illustrative construction example, the different aggregation rules had the biggest impact on the results. The results were different in the six calculation methods: when externally normalised scores were applied, house type A was superior in most of the calculations, but when internal normalisation was accomplished, house type C was superior. By using equal weights, similar results were obtained. None of the panellists intuitively considered A as the superior house type, but in some of the calculations, this was indeed the case. Furthermore, the results refer to the fact that the panellists completed the weighting on the basis of their general knowledge, without taking the features of different weight elicitation techniques into account.
Conclusions External normalisation provides information on a magnitude of impacts, and in some cases, external normalisation may be a more influential factor than weighting. Based on the results, it cannot be stated which different weight elicitation technique is the most suitable for LCA. However, the method should be selected based on the aims and purpose of the study. Moreover, the elicitation questions should be explained with care to experts so that they interpret the questions as intended.

Keywords Interval scale · Life cycle assessment · Multi-criteria decision analysis · Normalisation · Ratio scale · Weighting

Responsible editor: Göran Finnveden

T. Myllyviita (✉) · P. Leskinen · J. Seppälä
Consumption and Production Centre, Finnish Environment Institute, P.O. Box 140, 00260 Helsinki, Finland
e-mail: tanja.myllyviita@ymparisto.fi

1 Introduction

Life Cycle Assessment (LCA) is often performed to compare the environmental impacts of alternative products (e.g. Shen et al. 2010; Myllyviita et al. 2012). As a result of mandatory phases defined in the International Organization for Standardization (ISO) standards for LCA (ISO 2006), a decision-maker characterises scores of alternatives. Based on these characterised scores, the impacts of different products can be compared within one environmental impact, but the impacts cannot be compared across the impact categories. This is because characterised scores of different environmental impacts cannot be directly aggregated into a single score since they are measured in different units. It is not correct to aggregate, for instance, the CO_2 equivalents of

climate change and the Fe equivalents of metal depletion as such. Since there is no usual superior alternative (i.e. an alternative that has the smallest amount of environmental impacts with respect to all impact categories), value choices on the importance of environmental impacts are needed. The weighting of different impact categories is a commonly used approach when including value choices in LCA.

Value choices in the natural sciences have been widely discussed, especially since the meaning of sustainability is not consistent (Seager 2008; Bond et al. 2011). However, value choices in the natural sciences have also been denied, and the researcher's role as a “pure truth seeker” has been actively defended (Hacking 1999). However, subjective value choices are inevitable in LCA. For instance, defining system boundaries, selecting impact assessment method, how the allocation is actualised, and which impact categories are included during Life Cycle Impact Assessment (LCIA) (among many others) are influenced by values. Since weighting is readily identified as a subjective value choice, it has been a controversial topic in the field of LCA (Bengtsson and Steen 2000; Finnveden et al. 2002; Soares et al. 2006). Indeed, the ISO standard for LCA does not recommend weighting in comparative studies aimed at the general public (ISO 2006). Nevertheless, weighting has become a typical phase of LCA and even default weights are available. Some LCA methodologies recommend using equal weights (Jolliet et al. 2003), but using equal weights should not be considered as an objective selection. If equal weights are assumed, the impacts are considered equally important, which is also a value choice.

The weighting methods can be categorised as distance-to-target methods, monetary valuation methods and panel weighting methods. In distance-to-target methods, impact categories are evaluated on the basis of their distance from the current level to a future target level (Seppälä and Hämäläinen 2001), whereas monetary valuation methods assess the importance of the impact categories based on costs related to environmental consequences (Ahlroth et al. 2011). In this article, we focus on the panel weighting methods, which have been considered, along with monetary valuation methods, as the most promising approaches to weighting in LCA (Finnveden 1999). In the panel approach, experts or stakeholders are requested to weight impact categories using different elicitation procedures.

LCIA results can be actualised from midpoint or endpoint approaches (Bare et al. 2000). In the midpoint approach, the object of characterisation modelling is a “midpoint” in the cause-and-effect chain within each impact category (e.g. radiative forcing in climate change, H + release in acidification), whereas endpoint modelling refers to an assessment describing observable environmental end points, such as years of life lost. In addition, weighting can be completed from the midpoint or endpoint perspective. The problem with midpoint perspective weighting is that the panellists typically receive dozens of impact categories to be weighted, whereas there are only a few impact categories in the endpoint approaches. For example, the Eco-Indicator 99 consists of human health, ecosystem quality, and resources (Goedkoop and Spriensma 2000). However, the results of endpoint impact categories include more uncertainty compared to the results of midpoint impact categories, because the midpoint modelling can be conducted by relatively accurate modelling methods. In principle, weighting in the context of midpoint approaches can be carried out by similar Multi-Criteria Decision Analysis (MCDA) methods (Seppälä et al. 2002). It is only the contents of indicators within midpoint and endpoint categories that differ. In this paper, we focus on midpoint perspective weighting. In that approach, we consider different types of MCDA methods.

2 Theoretical background

2.1 Relationship between LCA and MCDA

MCDA is a family of methods that help decision makers identify and select a preferred alternative when faced with a complex decision-making problem characterised by multiple objectives (Von Winterfeldt and Edwards 1986). The benefits of using MCDA are a structured framework and the ability to deal with both qualitative and quantitative decision criteria with advanced calculation methods (Von Winterfeldt and Edwards 1986). Several authors (e.g. Miettinen and Hämäläinen 1997; Seppälä 1999, Seppälä et al. 2002; Azapagic and Clift 1998; Spengler et al. 1998; Soares et al. 2006; Seager and Linkov 2008; Jeswani et al. 2010; Myllyviita et al. 2012) have noted that MCDA can be applied to LCA. Most applications have been confined to weighting or to the use of LCIA results in the final decision context, although some authors (e.g. Seppälä et al. 2002) pointed out that MCDA can be applied in all stages of LCIA. A typical approach in LCA has been to use default weights for aggregating characterisation results into a single score. Such default weights can be found in the Eco-Indicator 99 method, for example (Goedkoop and Spriensma 2000). Soares et al. (2006) applied MCDA to the endpoint perspective, taking the scale and timeframe of impacts into account as well. The scale of impacts was also applied (with Analytic Hierarchy Process (AHP)) by Hermann et al. (2007). Outranking methods (Geldermann and Rentz 2005; Spengler et al. 1998) and AHP (Reza et al. 2011; Ong et al. 2001) have also been applied in LCA to aggregate single scores and to rank alternatives.

2.2 Interval and ratio scale

In MCDA, the criterion-specific sub-utility (or value or priority) functions are weighted in order to construct an overall score that describes the decision alternatives' multi-criteria values (Von Winterfeldt and Edwards 1986). Outranking methods, however,

indicate the degree of dominance of one alternative over another (Kangas et al. 2001). Depending on the characteristics of the sub-utility model, some weight elicitation techniques are based on the interval scale, while others are based on the ratio scale. The elicitation in the ratio and interval scale methods are different both in a theoretical and practical sense (Belton 1986). The ratio scale has a zero point, but at the interval scale, the zero point cannot be unambiguously defined. Multi-Attribute Value Theory (MAVT) (Von Winterfeldt and Edwards 1986) is one of the most commonly applied MCDA methods that is based on the interval scale. In MAVT, the most and the least preferred alternatives usually represent the extremities of the scale, and the other alternatives are compared to these and placed on the scale based on their distance from the extremities. Various weighting elicitation techniques can be included in MAVT. In SWING (Von Winterfeldt and Edwards 1986), the weights of decision criteria are obtained by considering the importance of the range of the attribute changing from its lowest level to the highest level. Simple Multi-Attribute Rating Technique (SMART) (Edwards 1971) and modifications of it called SMART using SWINGS and SMART Exploiting Ranks (Edwards and Barron 1994) can be based on the ratio scale, where one of the items is used as reference when evaluating the other items with respect to the reference. Another example of the MCDA method with the ratio scale is the AHP (Saaty 1971), which applies the ratio scale pairwise comparison of items. SMART can also be interpreted as a special case of AHP, where the minimum amount of pairwise comparisons is actualised. Originally, AHP was based on a 1–9 ratio scale, where 9 indicates that an item is nine times more preferred compared to the other item. However, numerous modifications on the original 1–9 scale have been also suggested (e.g. Alho et al. 2001; Leskinen 2001; Dong et al. 2008).

If weighting is accomplished with a method that is based on an interval scale, the weighting questions should be accomplished by referring to the impact assessment scores of decision alternatives (such as to the worst and the best alternatives in SWING). On the ratio scale, however, the weighting questions refer to criteria importance more generally, and criterion-specific utility changes are taken into account in the specification of the sub-utility functions. In both approaches, it is essential that the panellist understands the weighting questions exactly as is indented in the applied method.

2.3 External and internal normalisation in LCA

Normalisation, according to the ISO standard, is an optional phase in LCA, which can be applied after characterisation of impact assessment scores. In the LCA context, normalisation usually refers to external normalisation. External normalisation relates to a magnitude of impacts caused by a studied product system on a certain reference value (Finnveden et al. 2002; Norris 2001). External normalisation is often carried out to transform characterised LCIA scores to be more meaningful and to reveal the magnitude of impacts (e.g. Finnveden et al. 2002; Erlandsson and Lindfors 2003). External normalisation can be based on characterised scores caused by a certain geographical area or population. External normalisation factors have been compiled for European countries (Sleeswijk et al. 2008), Canada and the United States (Lautier et al. 2010). For instance, after external normalisation it could be stated that product A is responsible for 5 % of the annual climate change impacts of an average European citizen.

When MCDA is applied to aggregate single scores in LCA, internal normalisation becomes relevant. In internal normalisation, the reference values are not required, but the idea is to compare the decision alternative's value to other alternatives instead of external reference. With interval scale models in LCA, internal normalisation can be conducted by referring to the two extreme options within one criterion (Seppälä and Hämäläinen 2001). With ratio scale models, the relative utility scores of decision alternatives can be interpreted by scaling them to sum up to one (e.g. Saaty 1971). For instance, it can be stated that the climate change impacts of product A receives a score of 0.6, whereas the impacts of product B receives 0.4, therefore product A is 1.5 times more harmful than B in terms of climate change.

Norris (2001) has addressed the fact that external and internal normalisation is not congruent. In an illustrative example based on interval normalisation, Norris (2001) showed how the ranking of two alternatives changed, when a new alternative was included. The change in the rankings is caused by the rank reversal problem (e.g. Leskinen and Kangas 2005). Furthermore, Myllyviita et al. (2012) showed that relative impacts of alternatives assessed with LCA changed significantly when external normalisation was replaced with internal normalisation. Therefore, deciding which normalisation procedure is actualised, can have an impact on the results, especially in comparative studies.

2.4 Calculating a single score in LCIA

Decision makers who regularly use LCA to assist in decision making prefer to use single indexes since they seem to facilitate interpretation of the results (Mettier and Hofstetter 2004). In LCIAs, a typical calculation rule for a single score called total environmental impact (EI) is a simple model:

$$EI(a_j) = \sum_{i=1}^{n} w_i \cdot \frac{I_i(a_j)}{R_i} \tag{1}$$

where

$(EI)\,(a_j)$	Environmental impact of an alternative a_j
w_i	Weighting factor of impact category i
$I_i(a_j)$	Impact category indicator result of impact category i caused by alternative a_j

R_i reference value, i.e. indicator result of impact category i of the reference area

Equation (1) corresponds to the current LCIA methodology with external normalisation, although there is a wide range of competing LCIA methods. Popular methods such as Environmental Theme (Bauman and Rydberg 1994) and Eco-indicator 99 (Goedkoop and Spriensma 2000) are based on using Eq. (1), although they differ from each other in terms of the overall assessment philosophy.

In the case of internal normalisation and MAVT, where an interval scale is being assumed, Seppälä and Hämäläinen (2001) showed that total EI can be calculated by

$$EI(a_j) = \sum_{i=1}^{n} w_i \cdot \frac{I_i(a_j) - I_{i,\min}}{I_{i,\max} - I_{i,\min}} \tag{2}$$

where $I_{i,\min}=\min(I_i(a_1),\ldots I_i(a_m))$ and $I_{i,\max}=\max(I_i(a_1),\ldots I_i(a_m))$, where $_m$=number of alternatives

In other words, instead of external reference, alternatives' values are scaled with respect to other alternatives—the worst and the best alternatives in this case. According to MAVT, the sum of weighting factors should be adjusted to 1 ($\left(\sum_{i=1}^{n} w_i = 1\right)$) in both Eqs. (1) and (2). In Eq. (2), the weights are interpreted as the importance ratio of the contribution of the swing from worst to best in criterion i to the contribution of the swing from worst to best in criterion i. MAVT can assist in terms of how weighting factors should be assessed in the context of Eqs. (1) and (2). For example, weighing factors in Eq. (1) can reflect the damages caused by R_i and this feature can be taken into account in the determination of weights (see Seppälä 1999; Seppälä and Hämäläinen 2001; Seppälä et al. 2002).

Another possible way to calculate a total environmental impact in LCIA with internal normalisation is to utilise AHP

$$EI(a_j) = \sum_{i=1}^{n} w_i p(I_i(a_j)), \tag{3}$$

where w_i is obtained from pairwise comparison data of indicators' relative importance (see Saaty 1971; and e.g. Alho et al. 2001 for alternative estimation techniques). Also, here, the sum of weighting factors is adjusted to 1, (i.e. $\sum_{i=1}^{n} w_i = 1.$). In Eq. (3), $p(I_i(a_j)) = I_i(a_j) / \sum_{j=1}^{m} I_i(a_j)$. This means that the impact category indicator results caused by the production systems is internally scaled to sum up to 1, i.e. $\sum_{j=1}^{m} p(I_i(a_j)) = 1$. Moreover, the ratios $p(I_i(a_j))/p(I_i(a_j^{'}))$ can be obtained directly by ratios of alternative-specific impact assessment scores by assuming linear utility with respect to original impact assessments. Note that the analogous linearity assumption is also made with Eq. (2). Note also that if one is not willing to assume linearity, techniques exist to specify the shape of the sub-utility functions. In AHP (Eq. (3)), the weights represent the importance ratio of the two criteria with respect to the overall goal.

None of the methods mentioned above give an unambiguous solution to how the weighting process should be actualised or which technique should be applied in LCA. Several features of a weighting process can have an impact on the results. First, different methods may end up with different results (Bengtsson and Steen 2000). Secondly, if an MCDA method based on the interval scale is applied, it should be decided as to how impact assessment scores should be introduced to the panellists. External normalisation of the impact assessment scores may be necessary to make the scores of alternatives more meaningful for the panellists. It is also possible to assess the trade-offs between the reference values of different impact categories directly, without referring to the actual decision alternatives (Seppälä 1999). Here, the rationale would be to assume that impacts caused by a large area are easier to evaluate than the impacts caused by the alternatives. Finally, it is not clear how much and what kind of information the panellists should be given on the decision alternatives, and how they are influenced by the information provided.

The aim of this article is to describe how the weighting processes with different weighting elicitation techniques commonly used in MCDA are applied to LCA, and to compare the processes and the results. Impacts of internal and external normalisation in a comparative study are analysed in detail. As a case study, we present an illustrative construction project which analyses the environmental impacts of different types of houses with respect to three environmental impact categories (i.e. climate change, natural land-use change and metal depletion). Finally, based on the results, the problems related to weighting and aggregating single scores with different methods and some recommendations for weighting in LCA are given.

3 Methods

An illustrative construction project in the Nordic countries was assigned to apply weighting techniques. It was assumed that a neighbourhood with 20 detached houses in the Nordic countries was to be established and it was necessary to detect a house type with the smallest negative environmental impact. The three decision alternatives were as follows: A (brick house), B (wooden house) and C (log house). Three environmental impact categories—climate change, natural land transformation and metal depletion—were included. None of the houses is the best in all three environmental impact categories

(Table 1) therefore the superior house can only be detected via value choices. The impacts of the three houses with respect to the three environmental impact categories were characterised and externally normalised (proportion of Nordic citizens considered) (Sleeswijk et al. 2008) (Table 1).

The questionnaire with the six weighting exercises was compiled (Table 2) and Nordic practitioners and/or developers of LCA (who were mostly researchers), were engaged. The experts were identified by studying the list of participants at the 2011 NorLCA symposium and by reviewing journal articles related to LCA. All experts actualised the six weighting exercises in the same order (Table 2). In weighting exercise 1, the weights were acquired by applying the SMART ratio scale method. The question formatting is as follows "How much would you prefer to minimise metal depletion compared to climate change emissions released from Nordic countries?" The respondents were instructed to use the SMART ratio scale method: first, 100 points should be given to the environmental impact that he/she considered the most important, and others related to that (0–100 points). Aggregating single scores were based on Eq. (1).

In weighting exercise 2A, the same application of the SMART scale ratio method was used as in weighting exercise 1, but now, the question directly considered the construction case instead of the background reference. The panellists were given a detailed description of the construction project, three house types and the characterised scores of the environmental impacts of different houses (Table 1). The question formation was as follows: "How much would you prefer to minimise metal depletion in the Nordic countries compared to the climate change emissions when establishing a new neighbourhood with 20 detached houses in Nordic countries?" The aggregation in weighting exercise 2A was actualised in two ways, which are based on different assumptions on how the weights are interpreted. In the case of calculation method 2AI, characterised scores and Eq. (3) were applied. Note that SMART does not apply a full set (i.e. all of the items are compared to each other) of pairwise comparisons as in AHP, but SMART can be seen as a special case where items are compared with respect to one selected item. In addition to 2A1, question 2A was also interpreted through Eq. (1) by assuming that panellists were making assessments with respect to one characterisation unit. In other words, we assumed that normalisation factors are equal to one for every impact in Eq. (1).

AHP with the ratio scale was applied in weighting exercise 2B. The panellists were provided the same background information as in weighting exercise 2A. The pairwise comparisons were completed by giving the more important environmental impact 100 points and the other environmental impact points from 0–100 related to this (note that the scale in the original Saaty scale is 1–9). A full set of pairwise comparisons was made and Eq(3) was used to calculate the total environmental impact.

In weighting exercise 3, the SWING interval scale method was applied. Question formation was based on characterised scores of alternative house types (Table 1). The panellists were advised to assume that all three environmental impacts are at their highest (i.e. worst) level and to select the environmental impact that is the most important to change from the highest to the lowest (i.e. best) level. This environmental impact was given 100 points. Then, the panellists were asked to give points to the other two environmental impacts in a similar manner and to use the 100 points as a reference. While aggregating, single scores of alternative types of houses with respect to three environmental impacts were transformed into a 1–0 scale. The house with the smallest environmental impact with respect to a certain impact category received a value of 0, and the house with most environmental impact received a value of 1. A house with value between the two extremes, received a value between the 0–1 scale; this corresponds to Eq. (2).

Weighting exercise 4A was similar to weighting exercise 1, i.e. the weights were acquired by applying the SMART ratio scale method, but panellists were also provided *externally normalised* scores (Table 1). Calculation method 4A corresponds to Eq. (1).

Weighting exercise 4B used the SWING method with the interval scale (as in weighting exercise 3.) The difference between weighting exercise 3 and 4B is that the panellists in 4B are given *externally normalised* impact category scores (Table 1). The weights were acquired in a similar manner to weighting exercise 3, i.e. the panellists were encouraged to

Table 1 Environmental impacts (both characterised and normalised scores) of three types of houses (impacts/20 detached houses)

Environmental impacts of the three alternative types of houses			
	House type A	House type B	House type C
Climate change, characterised	678.000 CO_2 eq	444.000 CO_2 eq	137.160 CO_2 eq
Climate change, externally normalised	2.46E-06	1.61E-06	4.97E-07
Natural land transformation, characterised	26.7 m^2	53.4 m^2	63 m^2
Natural land transformation, externally normalised	6.71E-06	1.34E-05	1.58E-05
Metal depletion, characterised	926.220 Fe eq	836.250 Fe eq	836.250 Fe eq
Metal depletion, externally normalised	5.27E-05	4.76E-05	4.76E-05

Table 2 The six weighting exercises that were surveyed in the LCA case study. The experts of LCA were acquired to complete the six weighting exercises, that were using three different MCDA methods (SMART, SWING and AHP), different background information and scores

Descriptions of the six weighting exercises

Number of weighting exercise	Weight elicitation technique	Scale	Scores used in calculations	Background information provided for the panellists		
				Characterised scores	Externally normalised scores	Description of case study
1	SMART	Ratio	Externally normalised			
2A	SMART	Ratio	Characterised	X		X
2B	AHP	Ratio	Characterised	X		X
3	SWING	Interval	Characterised	X		X
4A	SMART	Ratio	Externally normalised		X	X
4B	SWING	Interval	Externally normalised		X	X

assess the importance of changing the values from the highest to the lowest value. In calculation method 4B, the scores were transformed to the interval scale and total environmental impacts were calculated according to Eq. (2). Note that now the external normalisation factors will be cancelled out, since they are used with internal normalisation, but the idea here was to see if the panellists thought that the weights should be different when impacts are normalised in the questionnaire.

After the six weighting exercises, the panellists were asked which of the three types of houses has the smallest environmental impacts with respect to the three environmental impact categories, and to give this house 100 points and to score the other two with respect to this (100–0 points). In this final weighting exercise, the panellists were not provided with any background information or scores, but they had to select the best alternative based on intuition, pre-questionnaire knowledge and information they had received during the previous weighting exercises. The questionnaire also included space for the respondents' comments and feedback.

The single scores of the three houses were aggregated by using weights from the six weighting exercises. The calculations were carried out separately for each panellist. The total environmental impacts of each house type within the six weighting exercises were also calculated by taking an average of the total environmental impact. Because some of the total environmental impact scores were calculated using the interval scale and others with the ratio scale, we transformed the calculations which were based on the ratio scale to the interval scale. The transformation was based on

$$REI\left(a_j\right) = \frac{\log\mathrm{EI}\left(a_j\right) - \min \log\mathrm{EI}\left(a_j\right)}{\max \log\mathrm{EI}\left(a_j\right) - \min \log\mathrm{EI}\left(a_j\right)} \tag{4}$$

where REI=rescaled total environmental impact and EI refers to original ratio scale results.

In Eq. (4), the idea is to first transform ratio scale values to the interval scale by log transformation (which transforms the multiplicative model to the additive) and then to rescale the outcome to a 0–1 scale.

4 Results

In total, 26 LCA experts gave their responses and feedback. Most of the respondents were from Finland (15), four were from Sweden, three from Norway and three from Denmark. Nine out of the 24 respondents considered themselves not only practitioners but also developers of LCA methodology.

House type C was superior (i.e. it had the lowest environmental impact) in most of the calculations (79 %) (Table 3) when all six weighting exercises are considered. House type A

Table 3 The superior type of house with respect to three environmental impacts (climate change, natural land transformation and metal depletion)

Calculation method	Percentage of calculations where house type was superior (total environmental impact)		
	House type A	House type B	House type C
1	70 % (0.29)	4 % (0.66)	27 % (0.68)
2AI	0 % (0.99)	41 % (0.74)	60 % (0.00)
2AII	0 % (1.00)	0 % (0.63)	100 % (0.00)
2B	0 % (0.99)	48 % (0.72)	52 % (0.00)
3	4 % (0.96)	4 % (0.54)	92 % (0.04)
4A	52 % (0.47)	12 % (0.51)	36 % (0.62)
4B	8% (0.92)	0 % (0.54)	92 % (0.05)

In the six weighting exercises, the results were remarkably different. The impacts were calculated using two methods: percentage of calculations where house type was superior and total environmental impacts calculated across the panellists (in parentheses). Note that the number in parentheses represents harmful impacts of a house, i.e. a house with a higher number is more harmful compared to a house with a lower number

had the lowest environmental impact in 19 % and house type B in 3 % of the calculations (Table 3). The different calculation methods produced different results, i.e. a different house had the lowest environmental impact in calculations based on different weighting exercises (Table 3). In the calculations based on weighting exercise 1, house type A was superior in most cases, but in the calculations based on weighting exercises 2A and 2B, house type A was not superior in any case. House type C was superior in the most calculations based on weighting exercises 2A, 2B, 3 and 4B. The total environmental impacts mostly support the same phenomenon: house type A has the smallest total environmental impact in calculation methods 1 and 4A, and house type C has the smallest total environmental impact in the other calculations (Table 3). When applying equal weights, house type A is superior in calculation methods 1 and 4A, and house type C is superior in all the other calculations (Table 4). The results are quite similar to the calculations based on the weights defined by the panellists in this paper (Table 3).

Based on their intuition, previous knowledge and information obtained from the six weighting exercises, half of the respondents considered house type C to be superior, and the other half considered house type B the superior type of house. None of the respondents considered house type A as the superior alternative.

The panellists had different opinions on which one of the six weighting exercise was the most meaningful. All of the six weighting exercises were considered the most meaningful by at least two panellists. Weighing exercise 2A was considered the most meaningful by seven of the panellists.

5 Discussion

Weighting in the context of LCA is not a thoroughly understood topic. Therefore, the aim of this paper was to further analyse how different weight elicitation techniques along with normalisation and background information influences a panel weighting process. The experts of LCA completed six weighting exercises, which were based on different weight elicitation techniques and background information. The weights defined by the panellists were used when calculating the environmental impacts of different houses. The impacts of three alternative house types appear to vary significantly when different calculation methods are applied (Table 3). Most of the differences in the results are caused by the aggregation method used in the calculations. Other factors that influence the results are the minor differences in the weights and behavioural aspects.

Three aggregation methods were used when calculating the impacts of the three house types. It can be seen that the results are similar in weighting exercise 1 and 4A (Eq. (1) used in calculations), 2AI, 2AII and 2B (Eq. (3) used) and finally 3 and 4B (Eq. (2) used). This outcome is notable, since all of the three calculation methods have a strong theoretical grounding, but there are no unambiguous instructions as to which one of them should be used in the context of LCA. Furthermore, there are several other MCDA methods and weighting techniques that could be used in LCA.

Using externally normalised scores may have a huge impact on the results compared to using characterised scores along with internal normalisation (Myllyviita et al. 2012; Shen et al. 2010). In this paper, house type A was superior compared to the other houses when externally normalised scores were applied (i.e. calculation methods 1 and 4A (Table 2)). Metal depletion is the source of most of the impacts (i.e. the score is larger than scores of land-use change and climate change) when externally normalised scores are used, but the differences in alternative house types with respect to metal depletion are minor, therefore metal depletion is not the most influential impact category when ranking the houses. Instead, when externally normalised scores are used, the land-use change will become the most influential impact category, since after external normalisation, the land-use change scores are larger than the scores of climate change and the differences with respect to land-use change are substantial (Table 1). In the calculations where scores are internally normalised (i.e. 2A, 2B and 3) the phenomenon described above does not emerge. Therefore, house type A was not the most environmentally friendly option in most cases when internal normalisation is actualised (Table 3).

The externally normalised scores provide information on the magnitude of impacts but the scores are so small that their interpretation may become difficult (Table 1). Indeed, some of the panellists mentioned in their feedback that the externally normalised scores were difficult to interpret. The respondents who considered weighting exercises 4A and 4B (where externally normalised scores were presented) as the most meaningful stated that it is important to know what the actual impacts are in terms of reference area. By intuition, climate change could be

Table 4 The superior type of house with respect to three environmental impacts (climate change, natural land transformation and metal depletion) when equal (i.e. weights for all three impacts are 0.333) weights are used. The results are similar to the results where weights provided by the panellists are used (Table 3)

Calculation method	Total environmental impact		
	House type A	House type B	House type C
1	0.00	0.37	1.00
2AI	1.00	0.85	0.00
2AII	1.00	0.55	0.00
2B	1.00	0.85	0.00
3	1.00	0.30	0.00
4A	0.00	0.37	1.00
4B	1.00	0.30	0.00

assumed to be the most relevant environmental impact category, since it has been a much discussed topic and countries have ambitious targets to cut their greenhouse gas emissions. Nevertheless, in the illustrative construction example described in this paper, metal depletion constitutes a larger proportion of the total metal depletion in the Nordic countries compared to climate change (Table 1). Consequently, one could consider in this specific case study that metal depletion is more important than climate change. Therefore, external normalisation reveals new aspects about the decision-making problem, which may be useful in the panel weighting process. However, the problem with external normalisation is that there are large uncertainties related to normalisation (Lautier et al. 2010; Heijungs et al. 2007), and normalisation factors related to less established impact categories may not be available.

Within the scope of this paper, behavioural aspects of individual panellists could not be thoroughly analysed. For instance, it is possible that the panellists preferred a certain house type, and therefore completed the weighting in a manner that supported their preference. In the questionnaire, nobody intuitively considered house type A superior, but in some of the calculations (especially those that were based on the externally normalised scores), house type A was superior. It is important that the panellists who complete the weighting do not let their personal preferences related to decision alternative impact on their weighting (assuming that their role is that of an expert). One possibility to avoid this would be to not reveal what the actual decision alternatives are to the panellists.

Based on the results of questionnaire, some evidence of possible misinterpretation can be detected. Different weighting exercises are based on different weight elicitation techniques, scale and background information, but still, the panellists completed the weighting in a remarkably similar manner in different weighting exercises. It can be seen that the average weights given by the 26 panellists in the six weighting exercises are similar (Fig. 1). Therefore, one relevant question, which cannot be answered based on the results of the questionnaire, is whether the panellists thoroughly understood the purpose of the different weighting exercises and the interpretation of the weights in each assessment task. However, since the interpretation of weights is difficult even for researchers in the field of MCDA (Choo et al. 1999), it is not surprising that this was difficult for the panellists as well, most of whom are not familiar with MCDA.

The primary purpose of the paper was to survey how different panel weighting methods, normalisation and background information provided for the panellists influence the results. Since external normalisation may be, at least in some cases, a more influential factor than weighting, it is recommended to actualise external normalisation before the weighting, since in some cases weighting may not be necessary. This is because the ranking of alternatives will not change after weighting, and using equal weights, for example, is admissible. If it seems that after external normalisation one of the well-established environmental impact categories constitutes most of the environmental load of the product, the weighting will not provide any new aspects, since the weights will not change the results. For instance, if it turns out that 95 % of the impacts of the products are caused by climate change, weighting in most cases will not change the results unless panellists consider climate change to be completely irrelevant (i.e. all panellists would give close to 0 points to climate change). On the contrary, if for instance it turns out that after external normalisation 95 % of environmental impacts of the product are caused by ecotoxicity, great care should be taken when analysing the results. First, it is likely that there might be errors in normalisation or impact assessment data or both, and reassessments should be made. In uncertain cases, weighting and internal normalisation, along with sensitivity and uncertainty analysis, would enhance the reliability of the results. In cases where after external normalisation it can be detected that environmental loads of alternative products are caused almost equally by several environmental impacts (i.e. there is no dominant impact category), the most environmentally friendly alternative can be detected only by weighting the impacts.

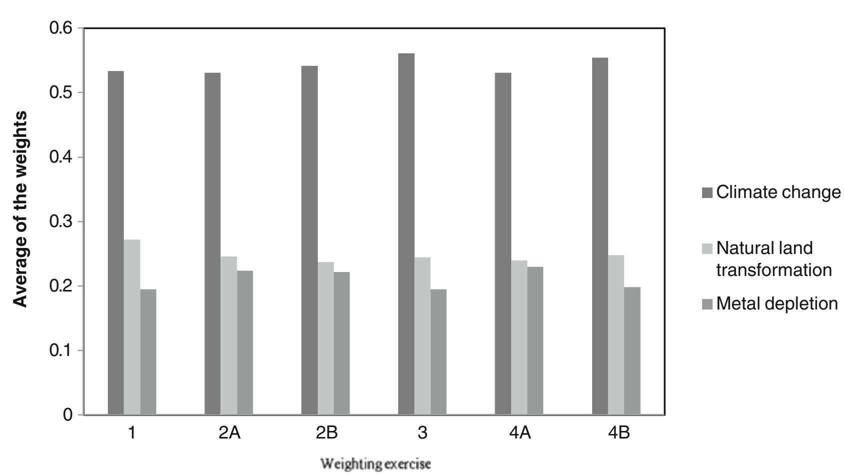

Fig. 1 The panellists completed the weighting in a moderately similar manner in the different weighting exercises, although the different weighting exercises were based on different methods, scales and background information

Based on the results of this paper, we cannot suggest one of the panel weighting methods as optimal. As discussed in previous sections of this paper, LCA is not free of value choices. Therefore, it is apparent that the aims and the purpose of the LCA have an impact on which panel weighting method should be applied and how. If LCA constitutes a larger region, such as a country, the weighting could be carried out in a more cursory manner, i.e. details or scores of the LCA are not needed (weighting exercise 1). If the case study is to be highlighted, the scores and details of the case study should be provided for the panellists. SWING is a promising technique for case studies like this, since the panellists are forced to closely evaluate and compare the scores of different decision alternatives. It cannot be stated based on the results of this paper whether the scores revealed to the panellists should be externally normalised. As discussed earlier in this paper, weighting can, in rare occasions, be avoided. External normalisation makes the magnitude of the impacts transparent, but it is not clear if the panellists are able to use this information when completing the weighting (Fig. 1). In this article, the panellists were researchers with knowledge of LCA, however, if for instance stakeholders and company managers are incorporated, another challenge is on how to make the weighting more understandable (Dahlbo et al. 2013). Endpoint perspective could be more understandable for the panellists; however, this has not been systematically tested. Focusing on weighting methods that are easy to interpret (and help the panellists to express their true judgements) should be favoured.

6 Conclusions

In situations where LCA is completed without normalisation and weighting, the problem is that characterised scores are aggregated and compared intuitively, resulting in interpretations that may not be transparent (Bengtsson and Steen 2000). According to the results of this paper, the intuitively selected superior alternative may not be the same as the results of an LCA analysis. Therefore, weighting could improve the transparency of LCA, since possible value choices are clearly stated. The most important conclusion based on the results of this paper is that results of comparative LCA studies are highly influenced by the applied aggregation rules. Another important topic is normalisation. External normalisation reveals the source of products' environmental impacts. If it is revealed via external normalisation that one of the impact categories constitutes most of the environmental load, the weighting may not be necessary, since the weights in most cases will not change the results.

Acknowledgments This study was funded by the Academy of Finland (decision number 127681). We truly appreciate the input of the 26 respondents who completed the weighting exercises and provided their valuable comments.

References

Ahlroth S, Nilsson M, Finnveden G, Hjelm O, Hochschorner E (2011) Weighting and valuation in selected environmental systems analysis tools—suggestions for further developments. J Clean Prod 19:145–156

Alho JM, Kolehmainen O, Leskinen P (2001) Regression methods for pairwise comparisons data. In: Schmoldt DL, Kangas J, Mendoza GA, Pesonen M (eds) The Analytic Hierarchy Process in natural resource and environmental decision making. Kluwer Academic Publishers, Dordrecht, pp 235–251

Azapagic A, Clift R (1998) Linear programming as a tool for Life Cycle Assessment. Int J Life Cycle Assess 3:305–316

Bare JC, Hofstetter P, Pennington DW, Udo de Haes HA (2000) Midpoints versus endpoints: the sacrifices and benefits. Int J Life Cycle Assess 8(65–735):319–326

Bauman H, Rydberg T (1994) Life cycle assessment: a comparison of three methods for impact analysis and evaluation. J Clean Prod 2:13–20

Belton V (1986) A comparison of the analytic hierarchy process and simple multi-attribute function. Eur J Opr Res 26:7–21

Bengtsson M, Steen B (2000) Weighting in LCA—approaches and applications. Environ Prog 19:101–109

Bond AJ, Dockerty T, Lovett A, Riche AB et al (2011) Learning how to deal with values, frames and governance in sustainability appraisal. Reg Stud 45:1157–1170

Choo EU, Schoner B, Wedley WC (1999) Interpretation of criteria weights in multicriteria decision making. Comput Ind Eng 37:527–541

Dahlbo H, Koskela S, Pihkola H, Nors M, Federley M, Seppälä J (2013) Comparison of different normalised LCIA results and their feasibility in communication. Int J Life Cycle Assess 18:850–860

Dong Y, Yinfeng X, Hongyi L, Min D (2008) A comparative study of the numerical scales and the prioritization methods in AHP. Eur J Opr Res 186:229–242

Edwards W (1971) Social utilities. Eng Econ Summer Symp Ser 6:119–129

Edwards W, Barron FH (1994) SMARTS and SMARTER: improved simple methods for multiattribute utility measurement. Organ Behav Hum Dec 60:306–325

Erlandsson M, Lindfors LG (2003) On the possibilities to apply the result from an LCA disclosed to public. Int J Life Cycle Assess 8:65–73

Finnveden G (1999) A critical review of operational valuation/weighting methods for Life Cycle Assessment. AFR-report 253, AFN (Swedish Waste Research Council). Swedish EPA, Stockholm

Finnveden G, Hofstetter P, Bare J et al (2002) Normalization, grouping and weighting in Life Cycle Impact Assessment. In: Udo de Haes HA, Jolliet O, Finnveden G (eds) Towards best practice in life cycle impact assessment – report of the second SETAC-Europe working group on life cycle assessment. Society of Environmental Toxicology and Chemistry (SETAC), Pensacola, pp 177–208

Geldermann J, Rentz O (2005) Multi-criteria analysis for technique assessment case study from industrial coating. J Ind Ecol 9:127–142

Goedkoop M, Spriensma R (2000) The Eco-indicator 99—a damage orientated method for Life Cycle Impact Assessment, 2th edn. Pre-Consultants N.V, Amersfoort

Hacking I (1999) The social construction of what? Harvard University Press, Cambridge and London

Heijungs R, Guinée J, Kleijn R, Rovers V (2007) Bias in normalization: causes, consequences, detection and remedies. Int J Life Cycle Assess 6:211–216

Hermann BG, Kroeze C, Jawjit W (2007) Assessing environmental performance by combining life cycle assessment, multi-criteria analysis and environmental performance indicators. J Clean Prod 15:1787–1796

ISO 14040 (2006) Environmental management. Life cycle assessment. Principles and framework. SFS-EN ISO 14040. Finnish Standards Association SFS

Jeswani HK, Azapagic A, Schepelmann P, Ritthoff M (2010) Options for broadening and deepening the LCA approaches. J Clean Prod 18: 120–127

Jolliet O, Margni M, Charles R et al (2003) IMPACT 2002+: a new life cycle impact assessment methodology. Int J Life Cycle Assess 8: 324–330

Kangas A, Kangas J, Pykäläinen J (2001) Outranking methods as tools in strategic natural resources planning. Silva Fenn 35:215–227

Lautier A, Rosenbaum RK, Margni M et al (2010) Development of normalization factors for Canada and the United States and comparison with European factors. Sci Total Environ 409:33–42

Leskinen P (2001) Statistical methods for measuring preferences. Dissertation, University of Joensuu

Leskinen P, Kangas J (2005) Rank reversals in multi-criteria decision analysis with statistical modelling of ratio-scale pairwise comparisons. J Oper Res Soc 56:855–861

Mettier TM, Hofstetter (2004) Survey insights into weighting environmental damages. Influence of context and group. J Indust Ecol 8:189–209

Miettinen P, Hämäläinen RP (1997) How to benefit from decision analysis in environmental life cycle assessment (LCA). Eur J Opr Res 102:279–294

Myllyviita T, Holma A, Antikainen R, Lähtinen K, Leskinen P (2012) Assessing environmental impacts of biomass production chains—application of life cycle assessment (LCA) and multi-criteria decision analysis (MCDA). J Clean Prod 29–30:238–245

Norris G (2001) The requirement for congruence in normalization. Int J Life Cycle Assess 6:85–88

Ong SK, Koh TH, Nee AYC (2001) Assessing the environmental impact of materials processing techniques using an analytical hierarchy process method. J Mater Process Tech 113:424–431

Reza B, Sadiq R, Hewage K (2011) Sustainability assessment of flooring systems in the city of Tehran: an AHP-based life cycle analysis. Constr Build Mater 25:2053–2066

Saaty TL (1971) A scaling method for priorities in hierarchical structures. J Math Psychol 15:234–281

Seager TP (2008) The sustainability spectrum and the sciences of sustainability. Bus Strat Env 17:444.453

Seager TP, Linkov I (2008) Coupling multicriteria decision analysis and life cycle assessment for nanomaterials. J Indust Ecol 12: 282–285

Seppälä J (1999) Decision analysis as a tool for life cycle impact assessment. The Finnish Environment 123. Finnish Environment Institute, Helsinki

Seppälä J, Hämäläinen RP (2001) On the meaning of the distance-to-target weighting method and normalisation in life cycle impact assessment. Int J Life Cycle Assess 6:211–218

Seppälä J, Basson L, Norris GA (2002) Decision analysis frameworks for life-cycle impact assessment. J Indust Ecol 5:45–68

Shen L, Worrell E, Patel MK (2010) Environmental impact assessment of man-made cellulose fibres. Resour Conserv Recy 55:260–274

Sleeswijk AW, van Oers L, Guinee JB, Struijs J, Huijbregts MAJ (2008) Normalisation in product life cycle assessment: an LCA of the global and European economic systems in the year 2000. Sci Total Environ 390:227–240

Soares SR, Toffoletto L, Deschenes L (2006) Development of weighting factors in the context of LCIA. J Clean Prod 14:649–660

Spengler T, Geldermann J, Hähre S, Sieverdingbeck A, Rentz O (1998) Development of a multiple criteria based decision support systems for environmental assessment of recycling measures in the iron and steel making industry. J Clean Prod 6:37–52

Von Winterfeldt D, Edwards W (1986) Decision analysis and behavioral research. Cambridge University Press, Cambridge